

LIMIT THEOREMS FOR THE MULTILEVEL SPLITTING ALGORITHM IN THE SIMULATION OF RARE EVENTS

Frédéric Cérou
 François LeGland

IRISA / INRIA
 Campus de Beaulieu
 35042 RENNES Cédex, FRANCE

Pierre Del Moral

Laboratoire J.-A. Dieudonné
 Université de Nice–Sophia Antipolis
 Parc Valrose
 06108 NICE Cedex 2, FRANCE

Pascal Lezaud

Laboratoire de Mathématiques Appliquées
 CENA / ENAC
 7 avenue Édouard Belin
 31055 TOULOUSE Cédex 4, FRANCE

ABSTRACT

In this article, a genetic-type algorithm based on interacting particle systems is presented, together with a genealogical model, for estimating a class of rare events arising for instance in telecommunication networks, nuclear engineering, etc. The distribution of a Markov process hitting a rare but critical set is represented in terms of a Feynman–Kac model in path space. Approximation results obtained previously for these models are applied here to estimate the probability of the rare events as well as the probability distribution of the critical trajectories.

1 INTRODUCTION

Let $X \equiv \{X(t), t \geq 0\}$ be a continuous-time strong Markov process taking values in some metric state space S . For a given Borel set $B \subset S$, define the hitting time

$$T_B = \inf\{t \geq 0 : X(t) \in B\},$$

as the first time when the process X hits B . The convention $\inf \emptyset = \infty$ is used throughout, so that $T_B = \infty$ if X never succeeds to reach B . Notice that the existence of left-hand limits insures that $X(T_B) \in \bar{B}$, where \bar{B} denotes the closure of B . In many applications, the set B is a (super) level set associated with a scalar measurable function ϕ defined on S , i.e.

$$B = \{x \in S : \phi(x) \geq \lambda_B\},$$

and if ϕ is upper semi-continuous, then $B = \bar{B}$ is a closed set, hence $X(T_B) \in B$.

Rare events are situations where most of the realizations of X never reach the target set B , which is often understood as a *critical* set. The corresponding rare event probabilities

$$\mathbb{P}[T_B \leq T], \tag{1}$$

$$\text{Law}(X(t), 0 \leq t \leq T_B \mid T_B \leq T),$$

are usually extremely difficult to evaluate. Here T is either a deterministic finite time, or a \mathbb{P} -almost surely finite stopping time, for instance the hitting time of a recurrent Borel set $R \subset S$, i.e. $T = T_R = \inf\{t \geq 0 : X(t) \in R\}$ and $\mathbb{P}[T_R < \infty] = 1$. The latter case covers the following two *dual* situations.

- The state space $S = A \cup R$ is decomposed into two disjoint sets A and R , the process X starts in A , and the objective is to estimate the probability that X hits the target set $B \subset A$ before it exits A (i.e. before it enters R).
- The state space $S = B \cup C$ is decomposed into two disjoint sets B and C , the process X represents a physical particle evolving in C , which contains a subset $R \subset C$ made of *hard obstacles*. The physical particle is killed as soon as it hits an obstacle, i.e. as soon as X enters R , and the objective is to estimate the probability that X escapes from C (i.e. hits the critical set B) without being killed.

Estimation of the rare event probabilities (1) is an important issue in a variety of scientific areas

- in telecommunication networks : probability that a buffer overflows at some node of the network, resulting in client rejection and packet loss,
- in nuclear engineering : probability that nuclear radiation escapes from the containment area of a nuclear plant,
- in air traffic management : probability that two airplanes get closer than a nominal separation distance (risk of conflict) or even hit one another (risk of collision),
- in environmental sciences : probability of occurrence of a major event (earthquake, flood, etc.),

and the design of fast, accurate and reliable estimation algorithms is a very active research topic. Several numerical methods have been proposed in the literature to estimate the probability of hitting a critical set, see Glasserman et al. (1999) which contains a precise review on these methods as well as a detailed list of references. For convenience, here is a brief description of the two main approaches.

The first approach is *importance sampling* and is based on changing the reference probability measure so that the rare event becomes less rare. This probabilistic method is often very sensitive to finding the right change of probability measure (this step is usually done using large deviations techniques). The second more physical approach is *multilevel splitting* and consists in decomposing the state space into a decreasing sequence of level sets the process X has to cross before it hits the critical target set. This decomposition step is usually based on a precise physical description of the evolution of the process between pairs of consecutive level sets. The next step is to introduce a system of particles evolving in this level-decomposed state space, in which each particle branches as soon as it enters into a smaller level set.

The purpose of the present article is to connect the multilevel splitting techniques with the branching and interacting particle systems approximations of Feynman–Kac distributions, studied in Del Moral and Miclo (2000), Del Moral and Miclo (2001) and Del Moral (2004). Further details on the application to rare events simulation can be found in Del Moral and Lezaud and in Krystul and Blom (2005). The objective here is twofold. First, a Feynman–Kac representation is proposed for the rare event probabilities (1). The general idea is to consider a discrete-time Markov chain in path space associated with excursions of the continuous-time Markov process from one level set to the next. This embedded Markov chain encodes all the information about how the continuous-time process crosses each level set before hitting finally the crit-

ical set. Based on this model, a natural mean-field type genetic particle approximation is introduced for the rare event probabilities (1). Moreover, and even more interestingly, the genealogical structure of the particle at each level can be used to estimate the probability distribution of the continuous-time process during its excursions towards the critical set.

Alternatively, an empirical method called *restart*, see Vill  n-Altamirano and Vill  n-Altamirano (1991), Vill  n-Altamirano and Vill  n-Altamirano (1994) and Tuffin and Trivedi (2000), can also be used to compute rare transient events and the probability of rare events in steady state, not only the probability to reach the target before coming back to a recurrent set. It is an advantage of the *restart* method with respect to multilevel splitting, but this method requires some mathematical approximations, and is not taken into account by the Feynman–Kac formalism. A further work could be an extension of the former particle scheme to cover the *restart* method as well.

The paper is organized as follows. Section 2 sets out the Feynman–Kac representation of the rare event probabilities (1). Section 3 introduces the genetic-type approximations based on interacting particle systems and their limit theorems. Finally, Section 5 deals with a numerical example, based on the Ornstein–Uhlenbeck process.

2 MULTILEVEL FEYNMAN–KAC FORMULAE

In practice, the process X starts in a Borel set $B_0 \subset S$ and before hitting the critical set B , he has to pass through a decreasing sequence of Borel level sets

$$B_0 \supset B_1 \supset \cdots \supset B_k \supset \cdots \supset B_n = B. \quad (2)$$

If $T = T_R$, then $B_0 = S \setminus R$, whereas if the final time T is deterministic, then $B_0 = S$. The decreasing sequence of level sets and their number depend on the problem at hand and in this paper, these design parameters are assumed to be given. An interesting issue would be to decide how many level sets should be considered, and how to choose these sets, but this question is out of the scope of this paper, see Section 3.3.

To capture the behavior of X between the different level sets $B_0 \supset B_1 \supset \cdots \supset B_k \supset \cdots \supset B_n = B$, introduce the discrete time embedded Markov chain $\mathcal{X} = \{X_k, k = 0, 1, \dots, n\}$ with values in the set

$$E = \bigcup_{t' \leq t''} \mathbb{D}([t', t''], S),$$

of trajectory excursions, induced from the continuous-time strong Markov process $X \equiv \{X(t), t \geq 0\}$ by the increasing

sequence

$$0 = T_0 \leq T_1 \leq \cdots T_k \leq \cdots \leq T_n = T_B ,$$

of stopping times, i.e. $\mathcal{X}_0 = X(0) \in E$ and

$$\mathcal{X}_k = (X(t), T_{k-1} \wedge T \leq t \leq T_k \wedge T) \in E ,$$

for any $k = 1, \dots, n$, where T_k represents the first time X reaches B_k , that is

$$T_k = \inf\{t \geq 0 : X(t) \in B_k\} ,$$

with the convention $\inf \emptyset = \infty$. Notice that

- if $T < T_{k-1}$, then $\mathcal{X}_k = X(T)$ and $X(T_k \wedge T) = X(T) \notin B_k$,
- if $T_{k-1} \leq T < T_k$, then $\mathcal{X}_k = (X(t), T_{k-1} \leq t \leq T)$ and $X(T_k \wedge T) = X(T) \notin B_k$,
- finally, if $T_k \leq T$, then $\mathcal{X}_k = (X(t), T_{k-1} \leq t \leq T_k)$ represents the trajectory excursion of X between the successive level sets B_{k-1} and B_k , and $X(T_k \wedge T) = X(T_k) \in B_k$,

and consequently $X(T_k \wedge T) \in B_k$ if and only if $T_k \leq T$. It follows that the stopping times can alternatively be defined by the inductive formula

$$T_k = \inf\{t \geq T_{k-1} : X(t) \in B_n\} ,$$

with the convention $\inf \emptyset = \infty$, so that $T_k > T$ if either $T_{k-1} > T$ or if starting in B_{k-1} at time T_{k-1} the process never reaches B_k before time T . Observe also that

$$(T_B \leq T) \Leftrightarrow (T_n \leq T) \Leftrightarrow (T_1 \leq T, \dots, T_n \leq T) .$$

By the strong Markov property, the stochastic sequence $(\mathcal{X}_0, \dots, \mathcal{X}_n)$ forms an E -valued Markov chain, with transition kernel

$$\mathcal{Q}_k(e, de') = \mathbb{P}[\mathcal{X}_k \in de' \mid \mathcal{X}_{k-1} = e] .$$

One way to check whether the path has succeeded to reach the k -th level is to consider the selection functions g_k on E defined for each trajectory excursion $e = (x(t), t' \leq t \leq t'') \in \mathbb{D}([t', t''], S)$ with $t' \leq t''$ by

$$g_k(e) = 1_{(\pi(e) \in B_k)} \quad \text{where} \quad \pi(e) = x(t'') .$$

With this notation, for each level k it holds

$$(T_k \leq T) \Leftrightarrow (T_1 \leq T, \dots, T_k \leq T)$$

$$\Leftrightarrow (g_1(\mathcal{X}_1) = 1, \dots, g_k(\mathcal{X}_k) = 1) ,$$

hence

$$1_{(T_k \leq T)} = \prod_{p=0}^k g_p(\mathcal{X}_p) ,$$

$$f(\mathcal{X}_k) \prod_{p=0}^k g_p(\mathcal{X}_p) = f(X(t), T_{k-1} \leq t \leq T_k) 1_{(T_k \leq T)} .$$

Introducing the nonnegative Feynman–Kac distribution γ_k defined by

$$\begin{aligned} \langle \gamma_k, f \rangle &= \mathbb{E}[f(\mathcal{X}_k) \prod_{p=0}^k g_p(\mathcal{X}_p)] \\ &= \mathbb{E}[f(X(t), T_{k-1} \leq t \leq T_k) 1_{(T_k \leq T)}] , \end{aligned}$$

for any bounded measurable function f defined on the set E of trajectory excursions, and in particular for $f \equiv 1$

$$\langle \gamma_k, 1 \rangle = \mathbb{P}[T_k \leq T] ,$$

the associated (normalized) probability distribution μ_k satisfies

$$\begin{aligned} \langle \mu_k, f \rangle &= \frac{\langle \gamma_k, f \rangle}{\langle \gamma_k, 1 \rangle} \\ &= \mathbb{E}[f(X(t), T_{k-1} \leq t \leq T_k) \mid T_k \leq T] , \end{aligned}$$

and in particular for any bounded measurable function ϕ defined on S

$$\langle \mu_k \circ \pi^{-1}, \phi \rangle = \langle \mu_k, \phi \circ \pi \rangle = \mathbb{E}[\phi(X(T_k)) \mid T_k \leq T] .$$

Similarly, introducing the nonnegative Feynman–Kac distribution $\gamma_{k|k-1}$ defined by

$$\begin{aligned} \langle \gamma_{k|k-1}, f \rangle &= \mathbb{E}[f(\mathcal{X}_k) \prod_{p=0}^{k-1} g_p(\mathcal{X}_p)] \\ &= \mathbb{E}[f(X(t), T_{k-1} \leq t \leq T_k \wedge T) 1_{(T_{k-1} \leq T)}] , \end{aligned}$$

for any bounded measurable function f defined on the set E of trajectory excursions, the associated (normalized) probability distribution $\mu_{k|k-1}$ satisfies

$$\begin{aligned} \langle \mu_{k|k-1}, f \rangle &= \frac{\langle \gamma_{k|k-1}, f \rangle}{\langle \gamma_{k|k-1}, 1 \rangle} \\ &= \mathbb{E}[f(X(t), T_{k-1} \leq t \leq T_k \wedge T) \mid T_{k-1} \leq T] , \end{aligned}$$

and in particular for $f = g_k$

$$\begin{aligned} p_k = \langle \mu_{k|k-1}, g_k \rangle &= \frac{\langle \gamma_{k|k-1}, g_k \rangle}{\langle \gamma_{k|k-1}, 1 \rangle} = \frac{\langle \gamma_k, 1 \rangle}{\langle \gamma_{k-1}, 1 \rangle} \\ &= \mathbb{P}[T_k \leq T \mid T_{k-1} \leq T] , \end{aligned}$$

hence

$$\begin{aligned} \mathbb{P}[T_k \leq T] = \langle \gamma_k, 1 \rangle &= \prod_{p=0}^k \langle \mu_{p|p-1}, g_p \rangle \\ &= \prod_{p=0}^k \mathbb{P}[T_p \leq T \mid T_{p-1} \leq T] . \end{aligned}$$

More generally, consider the path-space Markov chain $\mathcal{X}^\bullet = \{\mathcal{X}_k^\bullet, k = 0, 1, \dots, n\}$ defined by

$$\begin{aligned} \mathcal{X}_k^\bullet &= (\mathcal{X}_0, \dots, \mathcal{X}_k) \\ &= (X(0), (X(t), 0 \leq t \leq T_1 \wedge T), \dots \\ &\quad \dots, (X(t), T_{k-1} \wedge T \leq t \leq T_k \wedge T)) \\ &\equiv (X(t), 0 \leq t \leq T_k \wedge T) , \end{aligned}$$

with transition kernel

$$\begin{aligned} Q_k^\bullet(e_0, \dots, e_{k-1}, de'_0, \dots, de'_k) \\ &= \mathbb{P}[\mathcal{X}_k^\bullet \in (de'_0, \dots, de'_k) \mid \mathcal{X}_{k-1}^\bullet = (e_0, \dots, e_{k-1})] \\ &= \mathbb{P}[\mathcal{X}_0 \in de'_0, \dots, \mathcal{X}_k \in de'_k \mid \\ &\quad \mathcal{X}_0 = e_0, \dots, \mathcal{X}_{k-1} = e_{k-1}] \\ &= \delta_{e_0}(de'_0) \cdots \delta_{e_{k-1}}(de'_{k-1}) Q_k(e'_{k-1}, de'_k) , \end{aligned}$$

and consider the selection function defined by

$$g_k^\bullet(e_0, \dots, e_k) = g_k(e_k) = 1_{(\pi(e_k) \in B_k)} ,$$

for any (e_0, \dots, e_k) in the set E^{k+1} of concatenated trajectory excursions. Introducing the nonnegative Feynman–Kac

distribution γ_k^\bullet defined by

$$\begin{aligned} \langle \gamma_k^\bullet, f \rangle &= \mathbb{E}[f(\mathcal{X}_k^\bullet) \prod_{p=0}^k g_p(\mathcal{X}_p^\bullet)] \\ &= \mathbb{E}[f(X(t), 0 \leq t \leq T_k) 1_{(T_k \leq T)}] , \end{aligned}$$

for any bounded measurable function f defined on the set E^{k+1} of concatenated trajectory excursions, and in particular for $f \equiv 1$

$$\langle \gamma_k^\bullet, 1 \rangle = \langle \gamma_k, 1 \rangle = \mathbb{P}[T_k \leq T] ,$$

the associated (normalized) probability distribution μ_k^\bullet satisfies

$$\begin{aligned} \langle \mu_k^\bullet, f \rangle &= \frac{\langle \gamma_k^\bullet, f \rangle}{\langle \gamma_k^\bullet, 1 \rangle} \\ &= \mathbb{E}[f(X(t), 0 \leq t \leq T_k) \mid T_k \leq T] . \end{aligned}$$

The following Feynman–Kac representation holds for the rare event probabilities (1).

Theorem 1 (Feynman–Kac representations) For any level k

$$\mathbb{P}[T_k \leq T] = \langle \gamma_k, 1 \rangle ,$$

and

$$\mathbb{E}[\phi(X(T_k)) \mid T_k \leq T] = \langle \mu_k, \phi \circ \pi \rangle ,$$

for any bounded measurable function ϕ defined on S , and in addition

$$\mathbb{E}[f(X(t), 0 \leq t \leq T_k) \mid T_k \leq T] = \langle \mu_k^\bullet, f \rangle ,$$

for any bounded measurable function f defined on the set E^{k+1} of concatenated trajectory excursions.

The straightforward formula

$$\mathbb{P}[T_k \leq T] = \prod_{p=0}^k \mathbb{P}[T_p \leq T \mid T_{p-1} \leq T] ,$$

which shows how the very small probability of a rare event can be decomposed into the product of reasonably small but not too small conditional probabilities, each of which corresponding to the transition between two events, can be recovered from the the well-known identity

$$\langle \gamma_k, 1 \rangle = \prod_{p=0}^k \langle \mu_{p|p-1}, g_p \rangle ,$$

and will provide the basis for the efficient numerical approximation in terms of an interacting particle system. These conditional probabilities are not known in advance, and are learned by the algorithm as well.

3 MULTILEVEL GENETIC ALGORITHMS

In previous studies, see Del Moral and Miclo (2000), Del Moral and Miclo (2001) and Del Moral (2004), a collection of approximations based on branching and interacting particle systems have been designed to solve a general class of Feynman–Kac models. These particle techniques can also be used to solve the formulae presented in Theorem 1. This paper focuses on a simple mutation/selection genetic algorithm.

3.1 Nonsequential Particle Approximation

Recall that the evolution of the normalized (nonlinear) flow $\{\mu_k, k = 0, 1, \dots, n\}$ is described by the following diagram

$$\mu_{k-1} \mapsto \mu_{k|k-1} = \mu_{k-1} \mathcal{Q}_k \mapsto \mu_k = g_k \cdot \mu_{k|k-1},$$

with initial condition $\mu_0 = g_0 \cdot \eta_0$, where the notation \cdot denotes the projective product. The idea behind the particle approach is to look for an approximation

$$\mu_k \approx \mu_k^N = \sum_{i=1}^N w_k^i \delta_{\xi_k^i},$$

in the form of the weighted empirical probability distribution associated with the particle system $(\xi_k^i, w_k^i, i = 1, \dots, N)$, where N denotes the number of particles. The weights and positions of the particles are chosen in such a way that the evolution of the approximate sequence $\{\mu_k^N, k = 0, 1, \dots, n\}$ is described by the following diagram

$$\mu_{k-1}^N \mapsto \mu_{k|k-1}^N = S^N(\mu_{k-1}^N \mathcal{Q}_k) \mapsto \mu_k^N = g_k \cdot \mu_{k|k-1}^N,$$

with initial condition $\mu_0^N = g_0 \cdot S^N(\eta_0)$, where the notation $S^N(\mu)$ denotes the empirical probability distribution associated with an N -sample with common probability distribution μ . In practice, particles are trajectory excursions which

- are selected according to their respective weights $(w_{k-1}^i, i = 1, \dots, N)$ (selection step),
- move according to the Markov kernel \mathcal{Q}_k (mutation step),
- are weighted by evaluating the fitness function g_k (weighting step).

In other words, trajectory excursions that succeed to reach the next level set B_k before time T are selected and replicated, and the other trajectory excursions are terminated. Starting from the relation

$$\gamma_k = g_k(\gamma_{k-1} \mathcal{Q}_k) = g_k(\mu_{k-1} \mathcal{Q}_k) \langle \gamma_{k-1}, 1 \rangle, \quad (3)$$

and introducing the particle approximation

$$\gamma_k^N = g_k S^N(\mu_{k-1}^N \mathcal{Q}_k) \langle \gamma_{k-1}^N, 1 \rangle,$$

for the unnormalized (linear) flow, it is easily seen that

$$\frac{\gamma_k^N}{\langle \gamma_k^N, 1 \rangle} = g_k \cdot S^N(\mu_{k-1}^N \mathcal{Q}_k) = \mu_k^N.$$

However, the function g_k can only take the value 0 or 1, and even if $\langle \mu_{k|k-1}, g_k \rangle = \mathbb{P}[T_k \leq T \mid T_{k-1} \leq T] > 0$, it can happen that all the particles generated at the end of the mutation step fail to reach the next level set B_k before time T , in which case the evaluation of the function g_k returns the zero value for all the particles, and $\langle \mu_{k|k-1}^N, g_k \rangle = 0$: in such a situation, the particle system dies out and the algorithm cannot continue. A reinitialization procedure has been proposed and studied in Del Moral, Jacod, and Protter (2001), in which the particle system is generated afresh from an arbitrary restarting probability distribution ν . Alternatively, one could be interested by the behavior of the algorithm until the extinction time of the particle system, defined by

$$\tau^N = \inf\{k \geq 0 : \langle \mu_{k|k-1}^N, g_k \rangle = 0\}.$$

Under the assumption that $\langle \gamma_n, 1 \rangle > 0$, the probability $\mathbb{P}[\tau^N \leq n]$ that the algorithm cannot continue up to the time instant n goes to zero with exponential rate, see Theorem 7.4.1 in Del Moral (2004). On the *good set* $\{\tau^N > n\}$, it holds

$$\begin{aligned} \langle \gamma_n^N, 1 \rangle &= \prod_{k=0}^n \langle \mu_{k|k-1}^N, g_k \rangle \\ &= \prod_{k=0}^n \langle S^N(\mu_{k-1}^N \mathcal{Q}_k), g_k \rangle = \prod_{k=0}^n \frac{|I_k^N|}{N}, \end{aligned}$$

where

$$I_k^N = \{i = 1, \dots, N : g_k(\xi_k^i) = 1\},$$

denotes the set of successful particles within an N -sample with common probability distribution $\mu_{k-1}^N \mathcal{Q}_k$. In other words, the probability of a successful sequence is approximated as the product of the fraction of successful particles at each generation. Notice that the computational effort,

i.e. the number N of simulated particles at each generation, is *fixed* in advance, whereas the number $|I_k^N|$ of successful particles at the k -th generation is *random*.

The following results have been obtained for the non-sequential particle algorithm with a constant number N of particles : a nonasymptotic estimate, see Theorem 7.4.3 in Del Moral (2004), and a central limit theorem, see Section 9.4 of Del Moral (2004) for a slightly different algorithm.

Theorem 2 (\mathbb{L}^1 estimates) *For any level k*

$$\mathbb{E}[\langle \gamma_k^N, 1 \rangle 1_{(\tau^N > k)}] = \mathbb{P}[T_k \leq T],$$

i.e. the particle estimate $\langle \gamma_k^N, 1 \rangle$ is unbiased, and

$$\mathbb{E} \left| \frac{\langle \gamma_k^N, 1 \rangle}{\mathbb{P}[T_k \leq T]} 1_{(\tau^N > k)} - 1 \right| \leq \frac{c_k}{\sqrt{N}},$$

and

$$\sup_{\phi: \|\phi\|=1} \mathbb{E} |\langle \mu_k^N, \phi \circ \pi \rangle 1_{(\tau^N > k)} - \mathbb{E}[\phi(X(T_k)) | T_k \leq T]| \leq \frac{c_k}{\sqrt{N}}.$$

For any $k = 1, \dots, n$, any $l = 1, \dots, k$ and any bounded measurable function f defined on the set E of trajectory excursions, recall that

$$\begin{aligned} \mathcal{R}_{l+1:k} f(e) &= \mathcal{R}_{l+1} \cdots \mathcal{R}_k f(e) \\ &= \mathbb{E}[f(X_k) \prod_{p=l+1}^k g_p(X_p) | X_l = e] \\ &= \mathbb{E}[f(X(t), T_{k-1} \leq t \leq T_k) 1_{(T_k \leq T)} | \\ &\quad X(T_l) = \pi(e)], \end{aligned}$$

for any $e \in E$, where the nonnegative (unnormalized) kernel \mathcal{R}_p is defined by $\mathcal{R}_p(e, de') = Q_p(e, de') g_p(e')$ for any $p = 1, \dots, n$, and in particular for $f = \phi \circ \pi$

$$\begin{aligned} \mathcal{R}_{l+1:k}(\phi \circ \pi)(e) &= \mathbb{E}[\phi(X(T_k)) 1_{(T_k \leq T)} | X(T_l) = \pi(e)] \\ &= (R_{l+1:k} \phi)(\pi(e)), \end{aligned}$$

where

$$R_{l+1:k} \phi(x) = \mathbb{E}[\phi(X(T_k)) 1_{(T_k \leq T)} | X(T_l) = x],$$

for any $x \in S$, and in particular for $\phi \equiv 1$, let

$$R_{l+1:k} 1(x) = \mathbb{P}[T_k \leq T | X(T_l) = x].$$

Theorem 3 (CLT) *For any level k*

$$\sqrt{N} \left[\frac{\langle \gamma_k^N, 1 \rangle}{\mathbb{P}[T_k \leq T]} 1_{(\tau^N > k)} - 1 \right] \Longrightarrow \mathcal{N}(0, V_k^0),$$

with the following expression for the asymptotic variance

$$V_k^0 = \sum_{l=0}^k \left[\frac{\langle \mu_l \circ \pi^{-1}, |R_{l+1:k} 1|^2 \rangle}{p_l \langle \mu_l \circ \pi^{-1}, R_{l+1:k} 1 \rangle^2} - 1 \right],$$

and

$$\begin{aligned} \sqrt{N} \left[\langle \mu_k^N, \phi \circ \pi \rangle 1_{(\tau^N > k)} - \mathbb{E}[\phi(X(T_k)) | T_k \leq T] \right] &\Longrightarrow \mathcal{N}(0, v_k^0(\phi)), \end{aligned}$$

for any bounded measurable function ϕ defined on S , with the following expression for the asymptotic variance

$$v_k^0(\phi) = \sum_{l=0}^k \frac{\langle \mu_l \circ \pi^{-1}, |R_{l+1:k} [\phi - \langle \mu_k \circ \pi^{-1}, \phi \rangle]|^2 \rangle}{p_l \langle \mu_l \circ \pi^{-1}, R_{l+1:k} 1 \rangle^2}.$$

The impact of this asymptotic convergence theorem can be illustrated by choosing some particular test functions. For each $u > 0$, define the function $f(u)$ on the set E of trajectory excursions by setting for each $e = (x(t), t' \leq t \leq t'') \in \mathbb{D}([t', t''], S)$ with $t' \leq t''$,

$$f(u)(e) = 1_{(\ell(e) \leq u)} \quad \text{with} \quad \ell(e) = t'' - t'. \quad (4)$$

With this notation, the mapping $u \mapsto \mu_k(f(u))$ is the repartition function of the intertime $T_k - T_{k-1}$ between two consecutive level sets B_{k-1} and B_k , that is

$$\langle \mu_k, f(u) \rangle = \mathbb{P}[T_k - T_{k-1} \leq u | T_k \leq T].$$

The particle approximation of this quantity is the proportion of trajectory excursions that have managed to pass from B_{k-1} to B_k in time less than u .

3.2 Sequential Particle Algorithm

Alternatively, a sequential particle algorithm has been proposed in Le Gland and Oudjane (2004), which automatically keeps the particle system alive, i.e. which ensures its non-extinction. For any integer $H \geq 1$, and for any

$k = 0, 1, \dots, n$, define the random number

$$N_k^H = \inf\{N \geq 1 : |I_k^N| = H\},$$

of particles, where the random variables $\xi_k^1, \dots, \xi_k^i, \dots$ are i.i.d. with common probability distribution η_0 (for $k = 0$), and common probability distribution $\mu_{k-1}^H \mathcal{Q}_k$ (for $k = 1, \dots, n$), and where for any integer $N \geq 1$

$$I_k^N = \{i = 1, \dots, N : g_k(\xi_k^i) = 1\},$$

denotes the set of successful particles within the N -sample ξ_k^1, \dots, ξ_k^N .

The particle approximation $\{\mu_k^H, k = 0, 1, \dots, n\}$ is now parameterized by the number $H \geq 1$, and its evolution is described by the following diagram

$$\mu_{k-1}^H \mapsto \eta_k^H = S^{N_k^H}(\mu_{k-1}^H \mathcal{Q}_k) \mapsto \mu_k^H = g_k \cdot \eta_k^H,$$

with initial condition $\mu_0^H = g_0 \cdot S^{N_0^H}(\eta_0)$. Starting from (3) and introducing the particle approximation

$$\gamma_k^H = g_k S^{N_k^H}(\mu_{k-1}^H \mathcal{Q}_k) \langle \gamma_{k-1}^H, 1 \rangle,$$

for the unnormalized (linear) flow, it is easily seen that

$$\frac{\gamma_k^H}{\langle \gamma_k^H, 1 \rangle} = g_k \cdot S^{N_k^H}(\mu_{k-1}^H \mathcal{Q}_k) = \mu_k^H.$$

Clearly, $N_k^H \geq H$ and if $\langle \mu_{k-1}^H \mathcal{Q}_k, g_k \rangle > 0$ — a sufficient condition for which is

$$\widehat{g}_k(e) = \mathcal{Q}_k g_k(e)$$

$$= \mathbb{E}[g_k(\mathcal{X}_k) \mid \mathcal{X}_{k-1} = e]$$

$$= \mathbb{P}[T_k \leq T \mid X(T_{k-1}) = \pi(e)] > 0,$$

for any e in $\text{supp } \mu_{k-1}^H \subset B_{k-1}$ — then the random number N_k^H of particles is a.s. finite. Moreover

$$\begin{aligned} \langle \mu_{k|k-1}^H, g_k \rangle &= \langle S^{N_k^H}(\mu_{k-1}^H \mathcal{Q}_k), g_k \rangle \\ &= \frac{1}{N_k^H} \sum_{i=1}^{N_k^H} g_k(\xi_k^i) \geq \frac{H}{N_k^H} > 0, \end{aligned}$$

by construction, i.e. the particle system never dies out and the algorithm can always continue, and

$$\frac{H}{N_k^H} \longrightarrow p_k = \langle \mu_{k|k-1}, g_k \rangle > 0,$$

in probability as $H \uparrow \infty$, with rate $1/\sqrt{H}$. It holds

$$\begin{aligned} \langle \gamma_n^H, 1 \rangle &= \prod_{k=0}^n \langle \mu_{k|k-1}^H, g_k \rangle \\ &= \prod_{k=0}^n \langle S^{N_k^H}(\mu_{k-1}^H \mathcal{Q}_k), g_k \rangle = \prod_{k=0}^n \frac{H}{N_k^H}. \end{aligned}$$

Here again, the probability of a successful sequence is approximated as the product of the fraction of successful particles at each generation. In opposition to the nonsequential algorithm, notice that the number H of successful particles at each generation is *fixed* in advance, whereas the computational effort, i.e. the number N_k^H of simulated particles needed to get H successful particles exactly at the k -th generation, is *random*.

The following results have been obtained for the sequential particle algorithm with a random number of particles defined by the level $H > 0$: a nonasymptotic estimate, see Theorem 5.4 in Le Gland and Oudjane (2004) under a mixing assumption which is not needed here, and a central limit theorem, see Le Gland and Oudjane (2005).

Theorem 4 (\mathbb{L}^1 estimates) For any level k

$$\mathbb{E} \left| \frac{\langle \gamma_k^H, 1 \rangle}{\mathbb{P}[T_k \leq T]} - 1 \right| \leq \frac{c_k}{\sqrt{H}},$$

and

$$\begin{aligned} \sup_{\phi : \|\phi\|=1} \mathbb{E} | \langle \mu_k^H, \phi \circ \pi \rangle \\ - \mathbb{E}[\phi(X(T_k)) \mid T_k \leq T] | \leq \frac{c_k}{\sqrt{H}}. \end{aligned}$$

Theorem 5 (CLT) For any level k

$$\sqrt{H} \left[\frac{\langle \gamma_k^H, 1 \rangle}{\mathbb{P}[T_k \leq T]} - 1 \right] \Longrightarrow \mathcal{N}(0, V_k),$$

with the following expression for the asymptotic variance

$$V_k = \sum_{l=0}^k \left[\frac{\langle \mu_l \circ \pi^{-1}, |R_{l+1:k} 1|^2 \rangle}{\langle \mu_l \circ \pi^{-1}, R_{l+1:k} 1 \rangle^2} - p_l \right],$$

and

$$\sqrt{H}[\langle \mu_k^H, \phi \circ \pi \rangle$$

$$- \mathbb{E}[\phi(X(T_k)) \mid T_k \leq T] \implies \mathcal{N}(0, v_k(\phi)) ,$$

for any bounded measurable function ϕ defined on S , with the following expression for the asymptotic variance

$$v_k(\phi) = \sum_{l=0}^k \frac{\langle \mu_l \circ \pi^{-1}, |R_{l+1:k}[\phi - \langle \mu_k \circ \pi^{-1}, \phi \rangle]|^2 \rangle}{\langle \mu_l \circ \pi^{-1}, R_{l+1:k} 1 \rangle^2} .$$

Two different proofs can be given for this central limit theorem. A first proof follows the approach of Proposition 2.9 and Corollary 2.20 in Del Moral and Miclo (2000), and relies on an enumeration of all particles across generations with random sizes, and on a central limit theorem for triangular arrays of martingale increments. An alternate proof can also be given, which follows the approach of Theorem 4 in K  unsch (2003) by induction, and relies on the Anscombe central limit theorem for the sum of a random number of random variables, see Theorem I.3.1 in Gut (1988).

3.3 Comments on the Multilevel Decomposition

The multilevel genetic algorithms presented in this paper are usually sensitive to the choice of the decreasing sequence of Borel level sets (2). In practice, it is very difficult to choose these levels (and their number) in advance. In some applications, such as the evaluation of the risk of collision between two aircrafts, this decomposition is based on experience accumulated by practitioners, but this experience may be lacking in other applications.

Ideally, the multilevel decomposition should be chosen is such a way that the probability $\mathbb{P}[T_k \leq T \mid T_{k-1} \leq T]$ of hitting the k -th level set given that the $(k-1)$ -th level has already been reached, is independent of the index k . Based on this remark, adaptive strategies has been recently proposed in C  rou and Guyader, and limit theorems have been proved in the special case of a one-dimensional state space S , in which the levels are chosen on-line. Conceptually, it could even be argued that these new algorithms get rid of the notion of levels : indeed, they rely on the idea that it is enough to simulate N independent trajectories, to multiply a prescribed fraction $0 < p < 1$ (namely those pN trajectories which have managed to come closer to the critical set) of these trajectories, and to terminate the remaining $(1-p)N$ trajectories. In a certain sense, the worst among the best pN simulated trajectories defines a virtual level.

By construction, these algorithms guarantee a probability of success of p for the transition from one virtual

level to the next virtual level, whereas the number of virtual levels that should be crossed to finally hit the critical set is not known in advance, and is learned by the algorithm as well.

4 NUMERICAL APPLICATION TO THE ORNSTEIN-UHLENBECK PROCESS

In this section, a simple example is considered to illustrate how the algorithms presented above can be used to simulate rare events. Although this is clearly a toy model, it makes it possible to assess the method accuracy by computing some quantities for which an exact analytical expression is available. Moreover, the model is simple enough so that there is no numerical error due to discretization scheme.

The process X is the one-dimensional Ornstein-Uhlenbeck process, i.e. the solution of the SDE

$$dX(t) = -aX(t)dt + \sigma\sqrt{2a}dW(t) ,$$

where a and σ are strictly positive constants and W is a standard Brownian motion in \mathbb{R} . The recurrent set is $R = (-\infty, b^-]$, the target set is $B = [b^+, +\infty)$, and the process X starts at some x_0 with $b^- < x_0 < b^+$. It is clear that if b^+ is taken large enough, the probability to hit the target set can be made arbitrarily small.

Let τ denote the stopping time

$$\tau = \inf\{t > 0 : X(t) \notin (b^-, b^+)\} .$$

In order to check the method, the expected time $\mathbb{E}[\tau \mid X(\tau) = b^+]$ is computed using the multilevel genetic algorithm described in Section 3.1, and compared with the exact analytical expression. Indeed, it follows from Borodin and Salminen (1996) that

$$L(\alpha) = \mathbb{E}_{x_0}[e^{-\alpha\tau} 1_{(X(\tau) = b^+)}] = \frac{S(\frac{\alpha}{a}, \frac{x_0}{\sigma}, \frac{b^-}{\sigma})}{S(\frac{\alpha}{a}, \frac{b^+}{\sigma}, \frac{b^-}{\sigma})} ,$$

where S is a special function involving the gamma function and parabolic cylinder functions, which can be computed using the numerical method and the source code provided in Zhang and Jin (1996). Using the derivative of the Laplace transform yields

$$\mathbb{E}[\tau \mid X(\tau) = b^+] = -\frac{L'(0)}{\mathbb{P}[X(\tau) = b^+]} . \quad (5)$$

The probability in the denominator is given by

$$\mathbb{P}[X(\tau) = b^+] = \frac{u(x_0) - u(b^-)}{u(b^+) - u(b^-)} , \quad (6)$$

where the function u (the scale function of the process) is a primitive of $u'(x) = \exp\{\frac{1}{2}x^2/\sigma^2\}$. This function u is then easily computed using any standard numerical integration routine. An explicit expression of the derivative $L'(0)$ is more tricky to obtain, and a numerical approximation from a local rate of variation

$$L'(0) \simeq \frac{L(2\varepsilon) - L(\varepsilon)}{\varepsilon},$$

is used instead, where $\varepsilon > 0$ is chosen small enough.

The decreasing sequence (2) of Borel sets is defined here in terms of an increasing sequence $b^- = b_0 < b_1 < \dots < b_n = b^+$ of real numbers, as $B_k = [b_k, +\infty)$ for $k = 0, 1, \dots, n$. In this special case, the probability for a particle started at b_{k-1} to reach b_k can be fixed, and the number of levels and each level can be computed accordingly. If these probability are taken equal for all $k = 1, \dots, n$ to say p , then

$$n = \left\lfloor \frac{\log \mathbb{E}[\tau \mid X(\tau) = b^+]}{\log p} \right\rfloor.$$

Alternatively, the number n of levels can be fixed, and the probability p can be computed accordingly. Note that the probability of the N -particle cloud to be killed before hitting b^+ is $1 - [1 - (1 - p)^N]^{n+1}$, which can be small even with a small number N of particles when p is say larger than $\frac{1}{2}$. It follows that a good strategy is to make many runs of the algorithm with a small number N of particles, instead of only a few runs with a large number N of particles (on the same run, all the generated trajectories are obviously strongly correlated). The corresponding values b_k are easily computed for all $k = 1, \dots, n$, using expressions as in (6).

In Figure 1, the expectation $\mathbb{E}[\tau \mid X(\tau) = b^+]$ is displayed as a function of b^+ , with $b^- = 0$. The blue curve is the numerically computed theoretical value, and the red curve is the result of the multilevel genetic algorithm, with 880 runs of 8 particles each. The parameters of the Ornstein–Uhlenbeck process are $a = 0.1$, $\sigma \sqrt{2a} = 0.3$ and $x_0 = 0.1$. The largest value of b^+ was 4.0. This means that the probability for the process started at $x_0 = 0.1$ to hit the critical level is approximately 1.6460×10^{-8} , so there is no way of simulating critical trajectories by the naive approach.

ACKNOWLEDGMENTS

This work was partially supported by CNRS, under the projects *Méthodes Particulières en Filtrage Non-Linéaire* (project number 97–N23 / 0019, *Modélisation et Simulation Numérique* programme), *Chaînes de Markov Cachées et Filtrage Particulaire* (MathSTIC programme), and *Méthodes Particulières* (AS67, *DSTIC Action Spécifique* programme), and by the European Commission under the project *Dis-*

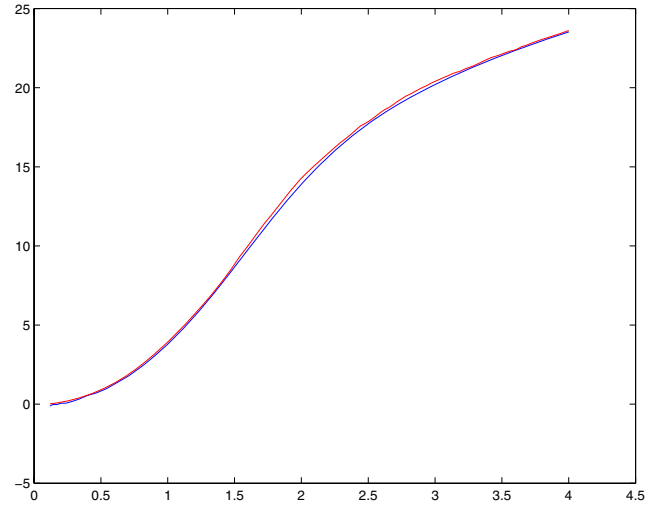


Figure 1: Theoretical and Monte–Carlo Mean Conditional Stopping Times

tributed Control and Stochastic Analysis of Hybrid Systems (HYBRIDGE) (project number IST–2001–32460, *Information Science Technology* programme).

REFERENCES

- Borodin, A. N., and P. Salminen. 1996. *Handbook of Brownian motion — Facts and formulae*. Probability and its Applications. Basel: Birkhäuser.
- Cérou, F., and A. Guyader, *An adaptive method for rare event estimation*. Preprint.
- Del Moral, P. 2004. *Feynman–Kac formulae. Genealogical and interacting particle systems with applications*. Probability and its Applications. New York: Springer–Verlag.
- Del Moral, P., J. Jacod, and P. Protter. 2001. The Monte Carlo method for filtering with discrete–time observations. *Probability Theory and Related Fields* 120 (3): 346–368.
- Del Moral, P., and P. Lezaud. Branching and interacting particle interpretation of rare event probabilities. In *Stochastic Hybrid Systems : Theory and Applications*, ed. H. Blom and J. Lygeros. Berlin: Springer–Verlag. To appear.
- Del Moral, P., and L. Miclo. 2000. Branching and interacting particle systems approximations of Feynman–Kac formulae with applications to nonlinear filtering. In *Séminaire de Probabilités XXXIV*, ed. J. Azéma, M. Émery, M. Ledoux, and M. Yor, Volume 1729 of *Lecture Notes in Mathematics*, 1–145. Berlin: Springer–Verlag.
- Del Moral, P., and L. Miclo. 2001, November. Genealogies and increasing propagations of chaos for Feynman–Kac

- and genetic models. *The Annals of Applied Probability* 11 (4): 1166–1198.
- Glasserman, P., P. Heidelberger, P. Shahabuddin, and T. Zazic. 1999, July–August. Multilevel splitting for estimating rare event probabilities. *Operations Research* 47 (4): 585–600.
- Gut, A. 1988. *Stopped random walks. Limit theorems and applications*. Applied Probability (A Series of the Applied Probability Trust). New York: Springer–Verlag.
- Krystul, J., and H. A. P. Blom. 2005, July. Sequential Monte Carlo simulation of rare event probability in stochastic hybrid systems. In *Proceedings of the 16th IFAC World Congress, Prague 2005*. IFAC. Paper Th–E07–TO.
- Künsch, H. R. 2003, January. Recursive Monte Carlo filters : Algorithms and theoretical analysis. Research Report 112, Seminar für Statistik, ETH, Zürich.
- Le Gland, F., and N. Oudjane. 2004, February. Stability and uniform approximation of nonlinear filters using the Hilbert metric, and application to particle filters. *The Annals of Applied Probability* 14 (1): 144–187.
- Le Gland, F., and N. Oudjane. 2005, September. A sequential algorithm that keeps the particle system alive. In *Proceedings of the 13th European Signal Processing Conference, Antalya 2005*. EURASIP.
- Tuffin, B., and K. S. Trivedi. 2000. Implementation of importance sampling techniques in stochastic Petri net package. In *TOOLS'00: Computer Performance Evaluation. Modelling Techniques and Tools*, ed. B. R. Harverkort, H. C. Bohnenkamp, and C. U. Smith, Number 1786 in Lecture Notes in Computer Science, 216–229. Berlin: Springer–Verlag.
- Villén-Altamirano, M., and J. Villén-Altamirano. 1991, June. RESTART : a method for accelerating rare event simulation. In *Queueing, Performance and Control in ATM: Proceedings of the 13rd International Teletraffic Congress, Copenhagen 1991*, ed. J. W. Cohen and C. D. Pack, Number 15 in North–Holland Studies in Telecommunications, 71–76. Amsterdam: North–Holland.
- Villén-Altamirano, M., and J. Villén-Altamirano. 1994, December. RESTART : a straightforward method for fast simulation of rare events. In *Proceedings of the 1994 Winter Simulation Conference, Orlando 1994*, ed. J. D. Tew, M. S. Manivannan, D. A. Sadowski, and A. F. Seila, 282–289.
- Zhang, S.-J., and J. Jin. 1996. *Computation of special functions*. New York: John Wiley & Sons.
- tion and Monte Carlo methods with interaction (particle methods). His e–mail address is [<fcerou@irisa.fr>](mailto:fcerou@irisa.fr).
- PIERRE DEL MORAL** is a professor of mathematics at Université de Nice–Sophia Antipolis. He has held visiting positions at Princeton University, in 2001, and at Purdue University, in 2003. His main research interests are Feynman–Kac formulas, Markov processes, interacting particle systems and idempotent probability calculus, on which topics he has organized or coorganized several workshops and has coordinated several national networks. He is the author of the monograph *Feynman–Kac formulae. Genealogical and interacting particle systems with applications* (2004). His e–mail address is [<delmoral@math.unice.fr>](mailto:delmoral@math.unice.fr).
- FRANÇOIS LE GLAND** is the head of a research group at IRISA / INRIA, Rennes. He has held visiting positions at the University of Maryland, in 1986, and at the University of Southern California, in 1993. His main research interests are statistics of stochastic processes, nonlinear filtering and Monte Carlo methods with interaction (particle methods), on which topics he has organized or coorganized several workshops and coordinated several national networks. His e–mail address is [<legland@irisa.fr>](mailto:legland@irisa.fr).
- PASCAL LEZAUD** is a researcher in the joint CENA / ENAC laboratory of applied mathematics, where he has been involved in the HYBRIDGE project funded by the European Commission, and in various statistical studies. His main research interests are collision risk models, Markov processes and rare event analysis. His e–mail address is [<lezaud@cena.fr>](mailto:lezaud@cena.fr).

AUTHOR BIOGRAPHIES

FRÉDÉRIC CÉROU is a researcher at IRISA / INRIA, Rennes. He has held a temporary position at Alcatel CRC, Marcoussis, and visiting positions at the University of Edinburgh, in 1995, and at Princeton University, in 1996 and 1999. His main research interests are rare event simula-